Running Head: NO CHILD LEFT BEHIND

No Child Left Behind and Tutoring in Reading and Mathematics: Impact of Supplemental Educational Services on Large Scale Assessment

Marco A. Muñoz

Jefferson County Public Schools

Florence Chang

Jefferson County Public Schools

Steven M. Ross

John Hopkins University

Paper presented at the annual meeting of the Consortium for Research in Educational Accountability and Teacher Evaluation (CREATE) 2011 National Evaluation Institute.

Abstract

The present quantitative study evaluated the effects of Supplemental Educational Services (SES), a federally mandated component of No Child Left Behind, on student achievement in reading and mathematics. SES provides free tutoring outside of school to disadvantaged students who attend Title I schools in their third year of failing to achieve Adequate Yearly Progress on state assessments. In the present study, data from the fifth year of SES implementation in a large urban school were analyzed to determine if the mostly small and nonsignificant effects obtained in prior years were stronger as the tutoring services acquired additional refinement and maturity. A matched treatment-control group design was employed, in which students who received SES tutoring in reading, mathematics, or both were matched to similar schoolmates who were eligible for SES services but did not participate. Results showed consistently positive but small, nonsignificant effect sizes for the outcome measures. Student achievement results and its implication for policy and practice are discussed.

Keywords: Supplementary Education: Achievement Tests; Educational Policy; Urban Schools; Federal Legislation; Program Effectiveness; Tutoring; Reading Skills; Mathematics Skills.

No Child Left Behind and Tutoring in Reading and Mathematics: Impact of Supplemental Educational Services on Large Scale Assessment

In 2001, the No Child Left Behind (NCLB, 2001) Act set forth legislation that aimed to have all children become proficient in reading and mathematics by 2013-2014. Congruent with NCLB's emphasis on schools using scientifically-based strategies and curriculum, the Education Sciences Reform Act (ESRA, 2002) states that it is critically important to identify "what works" in bringing all students to proficiency levels in core subjects. A central component of NCLB is the offering of Supplemental Educational Services (SES). SES provides free tutoring in reading and math outside of school to students who (a) attend Title I schools in the second year or more of school improvement (i.e., third year or more of failing to achieve Adequate Yearly Progress or AYP) and (b) are from lower-income backgrounds as defined by their participation in the free/reduced-price meal program. The SES providers and their services vary widely in their general characteristics including the number of tutoring hours provided per student, time of the year when tutoring occurs (e.g., before or after school hours, on weekends, or during the summer), and type of agency (e.g., community-based, faith-based, and private organizations). The school districts involved are required to reserve 20% of their Title I funds to support these services and school choice options.

Previous findings of SES implementation and achievement outcomes have been provided from studies conducted in Louisiana (Potter, Ross, Paek, Pribesh,

& Nunnery, 2006), Louisville (Muñoz, Potter, & Ross, 2008; Muñoz, Ross, & Neergaard, 2009), Los Angeles Unified School District (Rickles & Barnhart, 2007), Pittsburgh (Zimmer, Christina, Hamilton, & Prine, 2006), Tennessee (Ross et al., 2008), Milwaukeee (Heinrich, Meyer, & Whitten, 2010), and nationally across seven urban districts (Zimmer, Gill, Razquin, Booker, & Lockwood, 2007). Results have been mixed with some finding that higher participation in SES tutoring was associated with significant gains in reading and math (Zimmer et al., 2007), some finding only negligible or small impacts of SES tutoring on achievement (Rickles & Barnhart, 2007; Ross et al., 2008), and some finding differential effects on math and reading (Zimmer, Hamilton, & Christina, 2010). In a recent meta-analytic study, Chappell, Nunnery, Pribesh, and Hager (2011) again found small effect sizes for SES approximating +0.04 for mathematics and +0.02 in reading, and several provider characteristics associated with stronger impacts (e.g., district vs. external providers, well-trained tutors with four-year degrees, prescribed curriculum, one-on-one tutoring).

Most of the research evidence on SES, as reviewed above, synthesizes findings from multiple and diverse contexts across districts, states, and hundreds of providers (Stullich, Abrams, Eisner, & Lee, 2009). Not controlled is the degree of monitoring and quality control imposed at local levels to ensure that the tutoring is consistent with and connected to the curriculum, state assessment, class work, and student needs. The context for the present study was one urban school

district that has conducted its own SES evaluations for five years and used the data to work with providers and district staff to improve program quality and implementation fidelity. Thus, a reasonable hypothesis for this fifth-year study is that larger gains than realized in earlier years (Authors., 2008; Authors., 2009) and nationally (e.g., Chappell et al., 2011) would be found.

The Educational Value of Tutoring

One-to-one tutoring has long been regarded by many educators as the most powerful and adaptive method of instruction (Bloom, 1984). Not surprisingly, research has found that tutoring is an effective strategy to improve student academic outcomes. Elbaum, Vaughn, Hughes, and Moody (2000), for example, conducted a meta-analysis of reading programs using studies published between 1976 and 1998. Each of the studies included utilized adult tutoring interventions in an effort to improve the achievement of at-risk students. Results indicated that success can be achieved through effective tutoring with effect sizes as large as +.5. The Erlbaum et al. study supported the underlying approach of SES for using supplemental tutoring to increase the academic success of low-achieving students.

Research on early literacy programs conducted by Barley et al. (2002) also supports tutoring as an effective intervention for low-achieving or at-risk students. In reviewing prior research, the authors reached the following conclusions when describing effective tutoring programs: (a) education levels of the tutors can vary

as long as they are provided with appropriate tutor training, (b) tutoring sessions are monitored and adapted by program implementers, (c) the tutoring program has a strong guiding purpose (one that directs tutors in their decision making), (d) the program emphasizes the importance of diagnostic and prescriptive interaction, and (e) recruitment and retention of quality tutors are constantly a priority.

A review of supplemental programs conducted by Lauer et al. (2004; also see Lauer et al., 2006) extended Barley et al.'s work by including math programs in their analysis. Lauer et al. (2004) identified research and evaluation studies conducted after 1984 that addressed the effectiveness of programs delivered outside the regular school day for low-achieving or at-risk K−12 students. The synthesis resulted in statistically significant positive effects on student achievement in both reading and mathematics; the overall effect sizes ranged from .06 to .13 and from .09 to .17 for these subjects, respectively. These results suggest that tutoring programs can significantly increase the achievement of tutored students by an average of one-tenth of a standard deviation compared to students who do not participate in out-of-school tutoring programs. More recently, Chappell et al. (2011) reviewed two additional meta-analyses of tutoring effects, with results indicating larger effect sizes, specifically,.29 for reading and .60 for math (Cohen, Kulik, & Kulik, 1982); and .26 for reading and .27 for math (Ritter, Barnett, Denny, & Albin, 2009).

Results from other tutoring studies suggest that the most successful programs share common characteristics including (a) one-to-one tutoring structure, (b) systematic tutor training, and (c) continued program monitoring. In reviewing alternative interventions for struggling readers, Slavin, Lake, Davis, and Madden (2011) found that one-to-one tutoring was more effective than small group tutoring or computer-assisted instruction. Morris, Tyner, and Perney (2000) found that careful tutor training and formative evaluation of the tutoring sessions were key components to a quality early intervention reading program. Across studies, effective tutoring is characterized by personalized instruction that is adapted to individual learner needs from frequent diagnostic and prescriptive interchanges between tutor and tutee.

More prescriptive guidance in the form of five strategies for improving tutoring is provided from studies by Gordon and colleagues (Gordon, 2009; Gordon, Morgan, O'Malley, & Ponticell, 2007; Gordon, Morgan, Ponticell, & O'Malley, 2004): (1) use a diagnostic/develop- mental approach to help the tutor discover underlying student cognitive processing issues, such as learning disabilities (e.g., visual/auditory perceptual issues, attention-span limitations), (2) design and implement a highly structured tutoring program to help tutors implement more precise individualized tutoring, rather than generic "homework helper" or "drill-and-practice" tutoring that provides little assistance in improving student classroom achievement; (3) use highly trained tutors; (4) locate tutoring

sessions at sites likely to be accessible to students and their families; and, (5) take advantage of peer tutoring not only for increasing academic achievement but also for improving student motivation and self-efficacy.

In summary, previous research indicates that supplemental tutoring can be an effective way of adapting instruction to individual differences, particularly for students from at-risk backgrounds (Barley et. al., 2002; Gordon, 2009; Lauer et al., 2004, 2006; Slavin et al., 2011). In schools, teachers often face the challenge of differentiating their instruction in a classroom of students having a wide range of skill levels. Supplemental tutoring outside of the school allows students who are behind to potentially make greater academic gains and increase motivation for learning due to the individualized and additional instruction.

Purpose and Design of the Present Study

The NCLB (2001) legislation clearly and explicitly requires states to evaluate SES providers in terms of their effectiveness in raising student achievement. The present study of SES outcomes in a large urban district was designed to replicate prior comparison studies between SES students and matched control students on the state assessment in reading and mathematics (Authors, 2008; Authors, 2009). As districts near the 2013-14 NCLB deadline of all students reaching proficiency in reading and mathematics, the stakes are becoming higher for schools and for supplemental educational services to demonstrate success in increasing student achievement. As SES providers

become more experienced in providing services within particular districts and schools, it would be reasonable to expect the quality of tutoring to increase and concomitantly for students to show stronger gains.

The current study was conducted in a large urban district located in a metropolitan area. The district serves approximately 100,000 students in 150 schools, of which close to 20% were required to offer SES during the 2009-2010 school year. The majority of students are academically at-risk with over 60% participating in free and reduced lunch.

Important to the potential findings of the study, in this school year, the district staff used data collected by the district SES coordinator and from prior evaluation studies (e.g., Authors, 2008; Authors, 2009) to work systematically with service providers to improve program quality, better align the instruction with accountability measures, and ensure implementation fidelity. The major research question addressed concerned the degree to which the provision of SES services would raise achievement results for the students served both compared to control students and to prior evaluations of SES conducted in the same district. We addressed this question via a quantitative analysis of student-level scores on state-mandated tests in reading and mathematics.

The structure of SES programs prohibits the ability to conduct the most rigorous evaluation design, a randomized field trial. Parents whose children were eligible for SES services were provided information on the SES providers, and

then had to choose whether to enroll their children in services. Those who agreed then had to choose which provider would supply the services. Because a randomized design was precluded by parental choice, we adopted the next most rigorous option, a quasi-experimental design (Shadish, Cook, & Campbell, 2002) using closely matched program and control students with multiple student-level covariates.

An ancillary research question examined the impact of individual service providers. Since providers vary widely in their approach and intensity levels, it is important to examine whether some providers could serve as model programs for effectiveness. To evaluate the impact of different service providers, a pre-posttest-treatment-group-only design was used to examine and informally compare outcomes for individual service providers using statistical significance testing (i.e., paired-sample t-tests) and practical significance (i.e., effect sizes expressed as standardized mean differences).

Method

In the quasi-experimental design employed, students receiving SES services during the 2009-2010 school year were matched to demographically similar comparison students from the same schools who were also eligible for services but did not receive them. Given the variable sample sizes for providers, the initial analysis aggregated all the providers into one analysis with a follow-up pretest-posttest analysis to examine outcomes for individual providers.

Assessment of Achievement

The achievement measure employed was the Kentucky Core Content Test (KCCT) in Reading and Mathematics, with each test consisting of multiple-choice and open-response items. The test used for matching SES and control students on prior achievement and as the pretest covariate was the prior year's (2008-2009) KCCT test in the same subject. These criterion tests were group-administered and scored following standardized procedures (Kentucky Department of Education, 2009), and yield scale scores ranging from 1 to 80. A summary of the tests and analyses employed is provided in Table 1.

Insert Table 1 here

Matching Procedure

Data from the district database were used to identify the providers that tutored SES students during the 2009-2010 school year. The SES providers ranged from large national companies to local community-based organizations. Tutoring sessions were typically one hour after school, on two days per week, but varied in the methods of instruction. Some employed one-on-one or small-group instruction at a school or local site, while others tutored in the student's home or

online. The majority of programs lasted for several months, with most of the tutoring taking place in the second half of the school year.

The KCCT tests in Reading and Math are administered to all students in grades 3 through 8. In order to identify a matched-comparison group, the SES provider data file were combined with other databases containing student demographics and KCCT results for students in grades 4-8. Students in grade 3 were not included in the test because no previous state achievement baseline scores were available. Of the students who were eligible to receive SES services, 1,607 had both current (2009-2010) and prior-year (2008-2009) KCCT scores. Of these, 1,313 students actually received services, whereas 294 had applied for but did not participate in SES. Students received tutoring in (a) reading, (b) math, or (c) reading and math; separate analyses were conducted for each group of students.

Reading participants. The matched comparison group was based on student information from the 2008-2009 (prior school year) end-of-year database. Table 2 depicts the four variables used to match (a) students who received tutoring services (i.e., treatment group) and (b) students who did not receive tutoring services (i.e., comparison group). As shown, the matches were based on the prior year (2008-2009) KCCT test scores in Reading, gender, race, special education status, and limited English proficiency (LEP) status. All students included were active participants in the free or reduced-price lunch program. Chi-

square tests were conducted to examine the comparability of the treatment and comparison group on categorical variables. Results indicated that the groups were comparable with the exception of gender, with a greater percentage of males (61%) in the SES than control (48%) sample: [gender, χ (1, N = 691) = 11.13, p = .004; race, χ (1, N = 691) = 3.24, p = .07; special education, χ (1, N = 691) = .01, p = .96; LEP, χ (1, N = 691) = .37, p = .54]. An ANOVA conducted on the prior-year Reading achievement score was nonsignificant, F (1, 685) = .42, p = .52. As a result of the chi-square and ANOVA tests, we included the dummy-coded gender variable as a covariate in the posttest analysis. Although prior achievement did not differ between groups, it was employed as a second covariate to increase statistical power.

Insert Table 2 here

Mathematics participants. Table 3 presents the variables on which treatment and control students were matched for the analysis of mathematics achievement. All students included were active participants in the free or reduced-price lunch program. Chi-square tests for categorical variables were significant for gender [χ (1, N = 661) = .54, p = .46]; race [χ (1, N = 661) = 5.12, p = .02]; special education [χ (1, N = 661) = 4.07, p = .04]; and limited English proficient [χ (1, N = 661) = 5.52, p = .02]. As shown in Table 2, the SES sample

relative to the control sample had higher percentages of females and LEP students but lower percentages of minority students and special education students. An ANOVA performed on the prior-year KCCT mathematics test scores was also significant, F (1, 661) = 5.52, p = .02, showing an advantage for SES (M = 30.56) over control (M = 27.00) students. Based on the analyses, we included three dummy-coded variables (i.e., race, special education, and LEP) as well as prior-year KCCT scores as covariates in the posttest analysis.

Insert Table 3 here

Reading and mathematics participants. Table 4 shows the data for the matching variables for the SES sample that received tutoring in both subjects and the comparison group. All students were participants in the free or reduced-price lunch program. Chi-square tests showed a significant difference between groups in the percentage of students who were LEP (15.9% vs. 3.0% for SES and comparison, respectively); no significant differences were evidenced for gender, race, and special education status, [gender, χ (1, N = 862) = .94, p = .33; race, χ (1, N = 862) = .29, p = .59; special education, χ (1, N = 862) = 1.25, p = .26; limited English proficient, χ (1, N = 862) = 31.86, p = .001]. An ANOVA showed no difference in prior-year achievement in reading [F (1, 849) = .02, p = .89] or mathematics [F (1, 849) = .71, p = .40]. Based on these results, we included the

LEP variable as a covariate in the posttest analysis; prior-year test scores in the appropriate subject served as a second covariate.

Insert Table 4 here

Results

Using KCCT Reading and Mathematics test scores as dependent variables in grades 4-8, SES participants were compared across service providers to matched control students. Table 5 displays a summary of the results of the student achievement analyses.

Insert Table 5 here

KCCT Reading Test Scores

Tests of the homogeneity of regression indicated that assumptions were met and that usage of analysis of covariance (ANCOVA) was justified. The ANCOVA performed on KCCT Reading posttest scores yielded significance for gender [F (1,684) = 26.19, p = .001] and prior-year Reading score [F (1, 683) = 492.54, p = .001] as covariates. However, a non-significant Program effect in

Reading [F (1,683) = 0.41, p = .52] was found between SES and the comparison students. The adjusted effect size, while positive, was only about +.04.

A follow-up sub-group analysis of number of service hours was conducted, using gender and prior-year test scores as covariates, and keeping the original comparison group (n = 294). The service hour variable was derived using a median split (i.e., 27.25 hours) for the entire reading sample, resulting in a "high" service hour group that averaged 35.31 hrs, (n = 194, SD = 5.47, range = 27 to 43) and a "low" service hour group that averaged 20.94 hrs. (n = 194, SD = 8.79, range = 1 to 27.25). The ANCOVA performed on KCCT Reading by group (i.e., high hour, low hours, comparison group) yielded significance for the covariate gender [$F(1, 676) = 14.24$, $p < .001$] and the spring 2009 KCCT reading scores [$F(1, 676) = 484.59$, $p < .001$]. More importantly, an overall significant effect was not found for the high (M = 31.58, SD = 16.49) and low (M = 31.95, SD = 17.95) service hour groups and the comparison group [$F(2, 676) = .54$, $p = .58$]. As in the original analysis, comparison students (M = 32.14, SD = 16.66) did not differ from either group.

KCCT Mathematics Test Scores

Tests of homogeneity of regression confirmed that assumptions to justify ANCOVA were met for KCCT Mathematics. Accordingly, an ANCOVA using the baseline KCCT scores, special education, LEP, and race as covariates was conducted. Results indicated that prior achievement [F (1,650) = 783.13, p = .001]

and special education [F(1,650) = 19.84, p = .001] covariates were significant; the LEP [F (1,650) = .57, p = .45] and race [F (1,650) = .11, p = .75] covariates were nonsignificant. The analysis of the program effect, however, was nonsignificant [F (1,650) = 1.25, p = .26]. Positive effect sizes were found for both unadjusted and adjusted means (+.20 and +.06, respectively).

As performed for Reading, a sub-group analysis by number of service hours based on the median participation hours (i.e., 29.00) was conducted, using prior-year test scores and special education as covariates. The high service hour group (n = 176) averaged 34.45 hours (SD = 4.65, range = 29 to 46) and the low service hour group (n = 176) averaged 18.70 hours (SD = 9.64, range = 1 to 28). The ANCOVA results indicated that the KCCT mathematics [$F(1, 641) = 777.37$, $p < .001$] and special education [$F(1, 641) = 20.20$, $p < .001$] were significant covariates. More importantly, the Program effect again was not significant [$F(2, 641) = .72$, $p = .49$], regardless of dosage.

KCCT Reading and Mathematics Test Scores

A Multivariate Analysis of Variance (MANCOVA) was used for the Reading and Mathematics analysis. LEP, used as a covariate, was not significant for either subject, but prior-year achievement was significant ($p < .05$) in both. As shown in Table 5, the treatment comparison in Reading [F (1,849) = .11, p = .74] was not significant and associated with unadjusted and nonadjusted effect sizes close to zero. Nor was the comparison significant in Mathematics [F (1,849) =

2.11, p = .15]; yet, small positive effect sizes were found for both unadjusted and adjusted means (+.09 and +.11, respectively).

As performed for Reading and Mathematics, a sub-group analysis by number of service hours based on the median participation hours (i.e., 31.25) was conducted, using prior-year test scores and LEP as covariates. The high service hour group (n = 274) averaged 41.71 hours (SD = 22.12, range = 31.25 to 107.25) and the low service hour group (n = 274) averaged 23.01 hours (SD = 8.88, range = 1 to 31). The MANCOVA results indicated that the KCCT reading [$F(1, 837) = 200.24$, $p < .001$] and mathematics [$F(1, 837) = 403.46$, $p < .001$] were significant covariates. More importantly, the Program effect again was not significant for Reading [$F(2, 837) = .14$, $p = .87$] or Mathematics [$F(2, 837) = 1.28$, $p = .28$], regardless of dosage.

Dosage and Outcomes

The previous analyses utilized dosage (i.e., number of tutoring hours) in a median split. Analyses were conducted to examine the number of hours of tutoring received and outcomes utilizing correlational analyses. Number of hours of tutoring in reading was not related to growth in reading achievement, $r(959) = + .05$, $p > .05$. However, tutoring dosage in math was weakly but significantly correlated with growth in math achievement, $r(929) = +.07$, $p < .05$.

Service Provider Analyses

Dependent-sample t tests were conducted to assess the impact of specified service providers on Reading and Mathematics test scores. To ensure reasonable sample size and power for detecting statistical significance, analyses were conducted only for service providers having a total sample size equal to or greater than $n = 10$ in the target subject. Application of this criterion yielded 7 providers in Reading, 10 in Mathematics, and 10 in Reading/Mathematics.

In *Reading*, one of the seven providers was associated with statistically significant effects, Educate On-Line Learning ($n = 26$, *Mpre* = 27.88, *Mpost*= 33.54,) had a positive effect size of +.34, $t(1,25) = 2.23$, $p = .04$). In *Mathematics*, one of the 10 providers, Ivy League Tutor, was associated with a marginally significant gain: $t\ (1,11) = 1.88$, $p = .09$, ($n = 12$, *Mpre* = 38.00, *Mpost* = 44.50,) and a strong effect size of +.40. In *Reading/Mathematics*, none of the 10 providers was associated with statistically significant results.

The range of effect sizes among service providers (with n = 10 or above) was relatively high, extending from a -.35 to a +.40. Table 6 provides a summary of the 12 service providers that had positive effect sizes in reading, mathematics or both.

Insert Table 6 here

Discussion

Supplemental education services (SES) have been in place for almost a decade now. SES was originally formulated as a compromise on educational choice between liberal and conservative members of the House of Representatives (Steinberg, 2011). The shared, ostensive goal was to raise achievement of students attending underperforming Title I schools. Yet, research to date, including the present study, has shown (a) only weak or modest effects at raising student achievement (e.g., Heinrich et al., 2010, Ross et al., 2008) and that (b) the effects widely vary across providers (e.g., Chappell et al., 2011). Additional questions have been raised as to whether the family choice option results in the students who stand to benefit the most from SES being served to the same degree as their peers (Steinberg, 2011). Given the substantial resources being dedicated to SES, it is incumbent on providers and program supporters to demonstrate that the benefits obtained justify continuing to use the Title I funding for its support. For the particular district under study, the SES budget of over six million dollars per year potentially could be used to serve students in more beneficial ways.. SES expenses need to be analyzed in the context of budget cuts across states.

By comparison, research on individual tutoring programs tend to show positive benefits of both one-on-one and small-group tutoring, especially when provided by trained tutors (Green, Alderman, & Liechty, 2004; Ismail & Alexander, 2005; Slavin et al., 2011). Several reasons for the seemingly higher

success rate of these independent programs relative to SES can be proposed. One is that the tutoring programs examined in these studies generally are well-structured and implemented in standard ways across sites. SES, however, involves multiple programs employing varied tutoring organizations, curricula, and approaches. Second, the individual tutoring programs are usually implemented in restricted contexts (e.g., a few schools, community centers, or classrooms within a school) under supportive conditions where the sites independently sought the program or volunteered to use it. SES, on the other hand, is imposed on the district by federal requirements without necessarily having buy-in from schools. Third, the actual studies of individual tutoring programs tend to have more control over methods and instrumentation than is the case for SES. Because SES providers are large in number, scattered in locations used for tutoring, obtaining meaningful data on implementation quality and tutoring methods becomes highly challenging in district and state-wide studies. Also, in most SES studies, such as the present one, state assessments comprise the critical outcome measure in accord with federal expectations for judging provider effectiveness. Such high stakes assessments tend to be much less sensitive measures of immediate learning gains than would individually administered reading or mathematics inventories (Linn & Miller, 2005; Lipsey, 1990; Popham, 2009).

The school district participating in the present study had conducted several evaluations of SES and found the student achievement results to be disappointing (Authors., 2008; Authors., 2009). In an effort to improve program effects, the school district increased efforts to provide feedback and guidance to providers, and strengthen the connection tutoring activities to the state curriculum, regular classroom learning, usage of assessment data, and adaptation to student needs. A reasonable assumption was that these strategies would increase the quality of tutoring and impacts on KCCT scores. However, disappointingly, the overall effect sizes, computed from adjusted means, were nonsignificant and small, +.03, +.06, and +.02/+.11 in Reading, Mathematics, and Reading/Mathematics, respectively. Prior studies of interventions in Title I schools suggest realistic, meaningful effect sizes to range from $d = 0.10$ to $d = 0.20$ (Borman et al., 2003).

Based on the findings, the strategies employed by the district did not increase the impacts of SES on the KCCT assessments. However, the effectiveness of those strategies in improving the tutoring services in other ways cannot be ruled out. As noted above, the evaluation of SES, with its multiple providers and tutoring approaches, faces many methodological challenges (Ross, Paek, & McKay, 2008)). It is also noteworthy that the median tutoring dosage per student in reading or mathematics was only about 28 hours in total. That amount of tutoring would be roughly equivalent to adding only one week to the school year. Expecting such limited time to improve performance by low-achieving

students on state assessments (many of whom perform several years under grade level) simply may be unreasonable. In typical evaluation studies of regular tutoring programs (e.g., Slavin et al., 2011), the dosage was much higher. Even so, in the present study, SES students who received more hours of tutoring did not outperform either those who received less or comparison students. Correlational findings, however, indicated a weak but significant tendency for increased tutoring dosage in mathematics to be related to higher achievement. A second limiting factor may be poor alignment of the tutoring instruction with the core content covered on the state achievement tests; we wonder how this factor will continue to play a role as more states are moving toward the core national standards. Tutoring in and of itself does not necessarily lead to achievement gains. Approaches having specific characteristics, such as one-on-one ratios and certified, trained tutors, are most likely to impact student outcomes (Barley et. al., 2002; Lauer et al., 2004, Slavin et al., 2011).

The direction of the present results suggest greater potential of SES to impact math achievement rather than reading. This trend has also been found in past research (Zimmer, Hamilton, & Christina, 2010). Perhaps math is compartmentalized to the degree that a connection can be made more easily between the intervention (i.e., tutoring) and classroom instruction (Herman, 2006). For example, a student may learn during a tutoring session to solve a math problem from a different perspective which supplements the problem-solving

strategies taught by the regular teacher. On the other hand, in both mathematics and reading, there is the risk that SES tutors will teach content or learning strategies that interfere with the ways that students are taught in the regular classroom. Past work has indicated that it is important for practitioners to be involved in selecting, implementing, and evaluating programs (Slavin & Fashola, 1998).

Study Limitations

Interpretation of the present findings should be carefully interpreted due to the limitations of the design. First, due to the inability to employ a randomized experimental design, firm causal conclusions cannot be reached. Sampling representativeness is limited by parent choice in program enrollment and participation. It is not clear why some parents choose some providers over others and this may lead to some confounding variables in the provider groups, such as city location (e.g., parents from one neighborhood choosing one provider). Differences of this sort would compromise external validity, and can be particularly problematic for research on self-selection programs such as SES where less than 20% of those eligible for the services are likely to participate (Zimmer et al., 2007). In addition, the specific characteristics of the providers in the current study were not examined. Variables such as instructor-student ratios, coordination of instruction with core content standards, and dosage could help explain differences in provider impacts (Chappell et al., 2011).

Policy and Practice Implications

Supplemental tutoring is an instructional strategy with high potential and long tradition for assisting low-achieving students (Ismail & Alexander, 2005; Wasik, 1997). From a policy perspective, the goals of the NCLB legislation and SES could be on a collision course with expectations that far exceed what an after-school program of limited hours can reasonably accomplish. With research showing mixed results , there is no clear picture on the impact of SES on achievement. At best, it appears from the research that the true effect size of 30-40 hours per year (less than two weeks) of tutoring may only be 5 to 10 one-hundredths of a standard deviation. From a practitioner's perspective, it is difficult for school systems to ensure monitoring of the quality of SES providers. State policy may be the key to regulating how SES providers are certified and monitored to ensure quality programming. However, effective services may also be addressed by close coordination between the school systems, the service providers, and parents. District officials need to work with providers to align their strategies with classroom practices and state-required standards. It is also important for service providers to continuously monitor students via formative assessments and to receive feedback on their students' performance on the state-mandated assessments.

According to NCLB, states must remove providers from the approved list if they fail to increase student achievement for two consecutive years. However, it

does not appear that this policy has been implemented by many states or affected more than a handful of providers. The wide variability of both provider quality and student outcomes suggest that we need to continue to research the assumption that requiring districts to contract with external tutoring organizations will improve student achievement (Burch, Steinberg, & Donovan, 2007). Aside from determining what achievement gains are reasonable, it would also seem important to determine to what degree, if any, SES impacts other educational outcomes such as students' motivation, learning goals, self-efficacy, and interest in their education in general.

References

Authors (2008). Supplemental Educational Services as a consequence of the NCLB legislation: Evaluating its impact on student achievement in a large urban district. *Journal of Education for Students Placed At Risk, 13*(1), 1-25.

Authors (2009). Supplemental Educational Services as a component of No Child Behind: A mixed-method analysis of its impact on student achievement. *Planning and Changing, 40*(3&4), 135-159.

Barley, Z., Lauer, P. A., Arens, S. A., Apthorp, H. S., Englert, K. S., Snow, D., & Akiba, M. (2002). *Helping at-risk students meet standards: A synthesis of evidence-based classroom practices.* Aurora, CO: Mid-continent Research for Education and Learning (McREL).

Bloom, B. S. (1984). The search for methods of group instruction as effective as one-to-one tutoring. *Educational Leadership, 41*(8), 4–17.

Borman, G. D., Hewes, G., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, *73*(2), 125-230.

Burch, P., Steinberg, M., & Donovan, J. (2007). Supplemental Educational Services and NCLB: Policy assumptions, market practices, emerging issues. *Educational Evaluation and Policy Analysis, 29*(2), 115-133.

Chappell, S., Nunnery, J., Pribesh, S., & Hager, J. (2011). A meta-analysis of Supplemental Educational Services (SES) provider effects on student achievement. *Journal of Education for Students Placed at Risk, 16(1)*, 1-23.

Cohen, P., Kulik, J., & Kulik, C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*, 237–248.

Education Sciences Reform Act. (2002). P. L. No. 107-279.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S .W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, *92*(4), 605–619.

Gordon, E. E. (2009). 5 ways to improve tutoring programs. *Phi Delta Kappan, 90(6),* 440-445.

Gordon, E. E., Morgan, R. R., O'Malley, C. J., & Ponticell, J. (2007). *The tutoring revolution: Applying research for best practices, policy implications, and student achievement.* Lanham, MD: Rowman & Littlefield Education.

Gordon, E. E., Morgan, R. R., Ponticell, J., & O'Malley, C.J. (2004). Tutoring solutions for No Child Left Behind: Research, practice and policy implications. *NASSP Bulletin 88 (638)*, 56-68.

Green, S. K., Alderman, G., & Liechty, A. (2004). Peer tutoring, individualized intervention, and progress monitoring with at-risk second-grade readers. *Preventing School Failure, 49*(1), 11-17.

Heinrich, C. J., Meyer, R. H., & Whitten, G. (2010). Supplemental Education Services under No Child Left Behind: Who signs up, and what do they gain? *Educational Evaluation and Policy Analysis, 32 (2),* 273-298.

Herman, S. (2006). CSR model implementation from school stakeholder perspectives. *Journal of Education for Students Placed At Risk, 11*, 279-294.

Ismail, H. N., & Alexander, J. M. (2005). Learning within scripted and nonscripted peer-tutoring sessions: the Malaysian context. *The Journal of Educational Research, 99*(2), 66-77.

Kentucky Department of Education (2009). *Commonwealth Accountability Testing System: 2008-09 Technical report.* Frankfurt, KY: Measured Progress.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. (2004). *Effectiveness of out-of-school-time strategies in assisting low-achieving students in reading and mathematics: A research synthesis.* Aurora, CO: Mid-continent Research for Education and Learning (McREL).

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research, 76*, 275–313.

Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Pearson Education.

Lipsey, M. W. (1990). *Design Sensitivity: Statistical Power for Experimental Research.* Newbury Park, CA: Sage.

Morris, D., Tyner, B., & Perney, J. (2000). Early steps: Replicating the effects of a first-grade reading intervention program. *Journal of Educational Psychology, 92*(4), 681–693.

No Child Left Behind Act. (2001) Pub. L. No. 107-110.

Office of Research, Evaluation, and Accountability; Office of Extended Learning Opportunities (2008). *SES Tutoring Program Evaluation Report – Year 4.* Chicago, IL: Chicago Public Schools.

Popham, W. J. (2009). *Unlearned lessons: Six stumbling blocks to our schools' success.* Boston, MA: Harvard Education Press.

Potter, A., Ross, S. M., Paek, J., Pribesh, S. L., & Nunnery, J. A. (2006). *Supplemental Educational Services in the State of Louisiana: 2004-2005.* Center for Research in Educational Policy, University of Memphis, Memphis, TN.

Rickles, J. H., & Barnhart, M. K. (2007). *The impact of supplemental educational services on student achievement: 2005-06*. Los Angeles, CA: Los Angeles Unified School District Planning, Assessment, and Research Division Publication N0. 295.

Ritter, G. W., Barnett, J. H., Denny, G. S., & Albin, G. R. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research, 79(1),* 3-38.

Ross, S. M., Paek, J., & McKay, D. (2008). Is Supplemental Educational Services beneficial to students? Evaluation issues and perspectives from statewide studies. *ERS Spectrum*, *26*(1) 23-32.

Ross, S. M., Potter, A., Paek, J., McKay, D., Sanders, W., & Ashton, J. (2008). Implementation and outcomes of Supplemental Educational Services: The Tennessee state-wide evaluation study. *Journal of Education for Students Placed at Risk, 13* (1), 26-58.

Shadish,W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin Co.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best evidence synthesis. *Educational Research Review, 6,* 1-26.

Slavin, R. E., & Fashola, O. S. (1998). *Show me the evidence! Proven and promising programs for American schools*. Thousand Oaks, CA: Corwin Press, Inc.

Steinberg, M. P. (2011). Educational choice and student participation: The case of the Supplemental Educational Services provision in Chicago Public Schools. *Education Evaluation and Policy Analysis*, *33*(2), 159-182.

Stullich, S., Abrams, A., Eisner, E., & Lee, E. (2009). *Title I implementation—Update on recent evaluation findings.* Washington, DC: U.S. Department of Education Office of Planning, Evaluation and Policy Development.

Wasik, B. A. (1997). Volunteer tutoring programs: Do we know what works? *Phi Delta Kappan*, *70*, 283-287.

Zimmer, R., Christina, R., Hamilton, L., & Prine, D. W. (2006). *Evaluation of two out-of-school programs in Pittsburgh Public Schools: No Child Left Behind's Supplemental Educational Services and State of Pennsylvania's Educational Assistance Program*. Washington, DC: Santa Monica, CA: RAND.

Zimmer, R., Gill, B., Razquin, P., Booker, K., & Lockwood III, J. R. (2007). *State and local implementation of the No Child Left Behind Act: Volume I—Title I school choice, supplemental educational services, and student achievement.* Washington, DC: U.S. Department of Education, Office of

Planning, Evaluation and Policy Development, Policy and Programs Study Service.

Zimmer, R. Hamilton, L., & Christina, R. (2010). After-school tutoring in the context of No Child Left Behind: Effectiveness of two program in the Pittsburgh Public Schools. *Economics of Education Review, 29,* 18-28.

Table 1

Summary of Analyses Comparing SES to Control Students on Student Achievement

Outcome Measure	Year	Analysis	Covariate(s)
Grades 4-8 KCCT Reading	2009-2010	ANCOVA	Prior achievement on KCCT Reading; Gender
Grades 4-8 KCCT Mathematics	2009-2010	ANCOVA	Prior achievement on KCCT Mathematics; race; special education; and, LEP
Grades 4-8 KCCT Reading and Math	2009-2010	MANCOVA	Prior achievement on KCCT Reading and Mathematics; LEP

Note. KCCT = Kentucky Core Content Test (statewide assessment associated with accountability system); LEP = Limited English Proficient.

Table 2

Reading Treatment and Comparison Students Matched on Key Characteristics (N = 687)

	SES Students				Comparison Students			
	M	SD	n	%	M	SD	n	%
Prior-Year Test								
KCCT	30.63	17.91	393		31.51	17.28	294	
Gender								
Female			135	34.3			132	44.4
Male			242	61.4			143	48.1
Race								
Minority			265	67.3			219	73.7
Non-Minority			129	32.7			78	26.3
Special Education								
Yes			79	20.1			60	20.2
No			315	79.9			237	79.8
Limited English Proficient								
Yes			9	2.3			9	3.0
No			385	97.7			288	97.0

Note. Only students with complete testing data were included in the analysis. An aggregated matching procedure was utilized.

Table 3

Mathematics Treatment and Comparison Students Matched on Key Characteristics (N = 656)

Students	SES Students				Comparison			
	M	SD	n	%	M	SD	n	%
Prior-Year KCCT	30.56	20.59	362		27.00	19.15	294	
Gender								
Female			176	48.4			132	44.4
Male			163	44.8			143	48.1
Race								
Minority			242	66.5			219	73.7
Non-Minority			122	33.5			78	26.3
Special Education								
Yes			52	14.3			60	20.2
No			312	85.7			237	79.8
Limited English Proficient								
Yes			26	7.1			9	3.0
No			338	92.9			288	97.0

Note. Only students with complete testing data were included in the analysis. An aggregated matching procedure was utilized.

Table 4

Reading and Mathematics Treatment and Comparison Students Matched on Key Characteristics (N = 862)

Students	SES Students				Comparison			
	M	SD	n	%	M	SD	n	%
Prior-Year KCCT								
Reading	31.68	18.29	558		31.51	17.28	294	
Math	28.24	21.03	558		27.00	19.15	294	
Gender								
Female			281	49.7			132	44.4
Male			260	46.0			143	48.1
Race								
Minority			434	76.8			219	73.7
Non-Minority			131	23.2			78	26.3
Special Education								
Yes			133	23.5			60	20.2
No			432	76.5			237	79.8
Limited English Proficient								
Yes			90	15.9			9	3.0
No			475	84.1			288	97.0

Note. Only students with complete testing data were included in the analysis. An aggregated matching procedure was utilized.

Table 5

Unadjusted and Adjusted Means and Standard Deviations for SES and Matched-Control Comparisons on KCCT Scale Scores in 2009-2010

Comparison Group and Test	*N*	*M*	M_{adj}	*SD*	ES[a]	ES[b]
Reading						
SES	393	31.79	32.22	17.23	-0.02	+0.04
Control	294	32.14	31.58	16.66		
Mathematics						
SES	362	28.21	26.96	20.25	+0.20	+0.06
Control	294	24.28	25.83	18.74		
Reading and Math						
SES Reading	558	32.63	32.36	17.93	+0.03	-0.02
Control	294	32.14	32.66	16.66		
SES Math	558	26.08	25.57	19.98	+0.09	+0.11
Control	294	24.28	24.88	18.74		

Note. [a]Effect sizes were computed from the unadjusted means; [b]Effect sizes were computed from the adjusted means.

Table 6

Summary of Providers Associated with Positive Effects Sizes in Reading or Mathematics

Provider	Subject	Sample n	Effect Size
ATS Educational Consulting	Reading	10	+.03
	Math	10	+.13
Huntington Learning Center-West	Reading	110	+.10
	Math	10	+.06
Yes! All Students Can Learn	Reading	49	+.08
Club Z	Reading	127	+.10
Sylvan Learning	Math	89	+.12
Academic In-Home Tutoring	Reading	78	+.12
Better Grade Tutoring	Reading	64	+.13
A to Z In-Home Tutoring	Math	34	+.15
	Reading	14	+.17
SOAR Tutoring	Math	17	+.16
Educational Resources	Reading	14	+.29
Educate-On-Line Learning	Reading	26	+.34
Ivy League Tutor	Math	12	+.40